
Archives manuscrites littéraires : l'apport du numérique pour l'édition et la recherche scientifique

Aurèle Crasson



Édition électronique

URL : <http://journals.openedition.org/genesis/112>

DOI : 10.4000/genesis.112

ISSN : 2268-1590

Éditeur :

Presses universitaires de Paris Sorbonne (PUPS), Société internationale de génétique artistique littéraire et scientifique (SIGALES)

Édition imprimée

Date de publication : 20 juin 2010

Pagination : 43-47

ISBN : 978-2-84050-697-3

ISSN : 1167-5101

Référence électronique

Aurèle Crasson, « Archives manuscrites littéraires : l'apport du numérique pour l'édition et la recherche scientifique », *Genesis* [En ligne], 30 | 2010, mis en ligne le 30 mai 2012, consulté le 07 décembre 2020. URL : <http://journals.openedition.org/genesis/112> ; DOI : <https://doi.org/10.4000/genesis.112>

Tous droits réservés

Archives manuscrites littéraires : l'apport du numérique pour l'édition et la recherche scientifique

Aurèle Crasson

À défaut de s'entendre sur une définition univoque de l'édition électronique tant celle-ci recouvre de champs d'application, de ramifications entrecroisées avec d'autres pratiques de l'ordinateur et de l'Internet, il s'agira dans ce court article de cerner ce que peut apporter aujourd'hui le numérique à l'édition de manuscrits d'écrivains et à la transmission des genres d'œuvres littéraires.

On trouve aujourd'hui sur la toile de plus en plus d'éditions¹ qui démontrent l'intérêt de cette démarche. L'édition génétique de textes trouve par le biais du numérique une alternative à l'édition imprimée, notamment parce qu'elle permet de passer à une autre échelle documentaire. Mais, cette transposition dans une technologie encore émergente au regard du livre, n'a pas encore trouvé de forme qui lui permettrait de s'émanciper d'un modèle historique et d'affirmer sa pertinence.

La technologie a pris tant d'importance que l'on pourrait parfois constater un recul du projet éditorial lui-même : éditer des images de manuscrits, les connecter avec leurs transcriptions et les interprétations critiques qui en découlent dans la visée d'une œuvre cohérente. Ce qui pourrait paraître simple se confronte à un enjeu majeur : assumer l'hypertexte non seulement comme une technologie informatique mais aussi comme une construction intellectuelle, partie prenante de la structure de l'édition génétique.

L'enjeu de la numérisation se résume dans le fond à peu de choses, mais elles sont fondamentales : ajouter de l'« intelligence » à un document par le biais du balisage et de l'enrichissement de métadonnées, isoler des unités d'information que l'on saura utiliser pour structurer un ensemble documentaire et lui donner une multiplicité de représentations.

De l'objet palpable à sa transposition immatérielle

Il est banal de parler de « livre total » lorsque est évoqué le médium numérique. En effet, tout document créé sur celui-ci est potentiellement l'origine ou le relais d'un autre. Leur connexion est pourvoyeuse de construction de sens et de lecture supposée d'un ensemble logique. Pourtant, à l'heure où les projets d'éditions électroniques génétiques se multiplient, il est remarquable que si peu de documents soient conformes aux recommandations de principe. Force est de constater qu'aucun des langages de balisage, aussi poussés soient-ils, la *Text Encoding Initiative (TEI)* par exemple, qu'aucun des logiciels du marché, généralement inappropriés ou vite obsolètes, ne sont aptes à répondre indépendamment les uns des autres à toutes les spécificités d'un avant-texte. Il faut comprendre et faire admettre aux néophytes de l'informatique que le contenu d'un feuillet et son support matériel, le texte et son lieu d'ancrage dans la page, le graphisme et sa fonction de support de l'écriture constituent des « informations » au sens numérique du terme. Si ces données font système sur le plan sémiotique et en tout cas dans le contexte matériel du manuscrit, elles sont discrètes (désolidarisées) dans le contexte de leur numérisation. Ce sont des unités d'information qu'il convient de documenter très précisément pour permettre leur communication sur un ordinateur et donner la possibilité à celui qui les manipule, dans une perspective d'édition, de les utiliser pour toutes sortes de représentations.

1. Pour n'en citer que quelques-unes : édition des manuscrits de *Madame Bovary* : <<http://bovary.univ-rouen.fr/>> ; édition des manuscrits de Stendhal : <<http://stendhal.msh-alpes.fr/manuscrits/index.php>> ; édition du groupe de recherche sur Gabrielle Roy : <<http://digital.library.mcgill.ca/groy/xml/cahierpage.php?c=3&t=l&page=1>> ; édition des manuscrits de Nietzsche : <www.nietzschesource.org/>.

S'il n'a pas conscience qu'un document informatique n'est qu'une suite de 0 et de 1, un éditeur prisonnier du modèle du livre sera probablement conduit à penser l'édition électronique comme un ensemble de textes et d'illustrations porté par la logique d'une structure inamovible.

Alors que les premières éditions se développaient à partir d'éditeurs HTML, de traitements de textes, de bases de données ou de logiciels de traitement d'image cherchant à « reproduire » l'apparence des manuscrits, à retrouver des occurrences ou à redistribuer des données selon des logiques particulières à chacun, les développements informatiques créés pour l'occasion permettent bien d'autres opérations. Dès lors que les documents sont transformés en informations, toutes les possibilités de calcul et de procédure numériques sont envisageables, reléguant le fac-similé au rang d'une des nombreuses représentations d'un avant-texte.

La technologie numérique transforme l'idée que nous avons d'un document. Un manuscrit est communément perçu comme le support de travail d'un écrivain. Pour le généticien, ce sont donc sans doute les données textuelles qui attire le regard en priorité. Pour le conservateur c'est le support, la mémoire patrimoniale de l'œuvre qui compte. Pour le sémioticien tout élément (traits, taches, dessins, graphies, écritures) fait sens. Pour d'autres, ce sera l'image renvoyée, ce « tout » qui fait œuvre et œuvre d'art. Il s'agit tout à la fois de traces porteuses de sens et d'images qui renvoient à la « fabrique » de l'écrivain.

La numérisation, en transformant des objets matériels en « documents numériques », implique un changement radical de point de vue et cette notion d'information est bien plus complexe qu'il n'y paraît. Réduire un manuscrit à du texte ou à de l'image, c'est exclure la dimension chronologique inhérente à ce support, qui est déjà en acte à l'échelle d'un folio. La caractéristique du manuscrit n'est pas tant d'être un document hybride où scriptographies et images interfèrent, mais réside plutôt dans sa dimension doublement temporelle, tenant à la linéarité du texte et à la chronologie de son écriture. C'est cet aspect-là qui finalement est le plus difficile à gérer car la correction (processus de révision) est très fortement liée à la séquentialité du texte (processus de production). Pour le généticien, cette caractéristique – que les linguistes

représentent généralement par deux axes donnant les ordres syntagmatique et paradigmatique – est source de nombreuses hypothèses d'interprétation quant à l'établissement de la chronologie globale du manuscrit. C'est cette multiplicité qui rend difficile toute restitution, mais c'est justement ce que le traitement informatique est capable de représenter sous forme de parcours hypothétiques (construction logique de nœuds à nœuds).

S'intéresser aux informations, c'est aussi s'interroger sur les moyens de les rendre visibles à l'écran – tant, d'ailleurs, pour élaborer l'édition, que pour la consulter. Une édition électronique de genèse serait ainsi un ensemble de documents dont la description et l'annotation permettent de créer des « représentations » d'un avant-texte impossibles à matérialiser dans une autre forme que numérique ; elle est donc le résultat de la dématérialisation du support analogique qui lui permet de s'actualiser en représentations virtuelles. Son aboutissement se constitue d'images décrites vectoriellement, de transcriptions textuelles enrichies d'une base de connaissance, et de relations entre ces ressources qui servent à décrire l'articulation entre des informations de nature différente. Ce que le lecteur découvre, ce sont des interfaces (des vues qui s'enchaînent par des actions de souris ou de clavier) qui ont interprété ces relations tout en rendant transparente la structure.

Structure

L'édition imprimée reste un modèle de référence, mais contrairement à l'éditeur papier qui construit un objet spécifique dans une catégorie éditoriale désignée, le numérique permet d'en élaborer plusieurs à partir de mêmes ressources. Ce sont l'organisation des niveaux de traitement et la circulation dans ces informations qui déterminent le type d'édition électronique. C'est aussi ce qui fonde l'hypertexte, qui se définit tantôt comme le produit d'une construction intellectuelle, tantôt comme une technologie autorisant la mise en relation de ressources et la constitution de parcours.

Qu'est-ce que la structure d'une édition de genèse ? Une sorte de trame dans laquelle on dispose des jalons facilitant l'accès à l'information, les interconnexions et voies d'échange.

On dit parfois qu'un avant-texte est lui-même un hypertexte. Il est intrinsèquement composé d'objets hétéroclites tangibles – ou interprétés car les traces ne sont pas toujours attestées (l'exactitude du temps des corrections notamment) – qui renvoient à une cohésion d'ensemble intimement liée à l'idiosyncrasie de l'écrivain. Le rôle du généticien est de faire émerger cet hypertexte – que je dirais « latent » – en actualisant des relations elles-mêmes constructrices de parcours (chronologiques, d'hypothèses d'interprétation, etc.).

Dans un contexte de recherche scientifique, la numérisation se justifie par l'idée qu'un document sera plus facile à consulter et à manipuler. L'édition électronique se présente comme un processus structurel tendant à chaque fois vers plus de seuils d'abstraction. Ces niveaux se rapportent à l'image (le fac-similé numérique), sa dissociation en éléments scripto-graphiques², textuels et topographiques.

Structurer un manuscrit suppose des partis pris. Je considère le feuillet comme un ensemble de zones connexes – que j'appelle « régions » par analogie géographique – pouvant parfois se superposer et qui « encapsulent » des éléments textuels, graphiques, métatextuels ou d'autres feuillets, des paperoles par exemple, eux-mêmes contenant d'autres régions. Ces unités visuelles – qui ne sont que des coordonnées topographiques – constituent une possible base structurelle pour une édition. Elles permettent par exemple de cartographier les temporalités du manuscrit à travers la relation qui unit les zones graphiques à leur contenu sémantique.

L'organisation du tracé des villes est très comparable à la conception d'une édition génétique dans le sens où il s'agit de faire emprunter à un utilisateur des parcours. Une structure qui confine à la fermeture d'un modèle sur lui-même est révélatrice d'une mauvaise approche de l'articulation des objets car elle supprime toute possibilité d'inattendu. D'un point de vue urbanistique, le concept des villes nouvelles né au milieu des années soixante et qui se voulait exemplaire (notamment pour le cadre de vie et l'économie) a, contrairement à l'espoir suscité, abouti à une impasse. Les éditions électroniques génétiques actuelles pâtissent de la même manière de modèles figés, ce qui est totalement contradictoire avec la technologie

numérique et avec le principe même de la genèse, qui n'est jamais qu'une interprétation ouverte à d'autres dans un tissu quasi organique. Même si la technologie permet de juxtaposer toutes sortes d'informations, le fait de séparer les interfaces de construction de la genèse (qui nécessitent par exemple des outils de transcription, de codage ou d'exploitation vectorielle) de celles de visualisation (qui nécessitent des outils de circulation et de cartographie) devrait inciter l'éditeur à s'affranchir de toute modélisation trop contraignante et à laisser au lecteur la possibilité de reconstituer un espace complexe en fonction de ses propres hypothèses.

Codage

Le codage sémantique est une opération très délicate, moins du point de vue de sa mise en œuvre que du choix de concepts pouvant faire consensus. Chaque chercheur est amené à créer des concepts et souhaiterait qu'ils soient codés informatiquement afin de pouvoir lancer des requêtes à partir de critères choisis.

Est-il possible pour autant de *tout* décrire pour être sûr que *rien* n'échappe à aucun spécialiste du manuscrit ? Probablement non ; d'une part, nul ne saurait embrasser cette totalité sémantique et conceptuelle et par ailleurs, cette surcharge d'informations ne conduirait pas à de meilleurs retours sur le travail savant. Comme toute archive est un puits sans fonds, toute surinformation est génératrice de « bruits ». Plus on précise l'information sur des concepts non génériques, plus on referme le manuscrit sur sa complexité. Décider en revanche des objets qui font consensus permet de parler de la même chose mais à travers des présupposés différents et des descriptions fouillées. La terminologie est donc primordiale même si elle donne lieu à des variantes ou des déclinaisons. L'exemple des ajouts et des suppressions est particulièrement remarquable. Un mot biffé c'est de l'écriture sur laquelle vient se superposer une trace graphique que d'un commun accord nous

2. Le niveau scripto-graphique, qui repose sur l'unité du feuillet, constitue un pivot dans la mesure où il lie les informations sémantiques à la topographie de l'image.

appelons « rature », « biffure », « rayure » ou bien d'autres synonymes moins en usage. Chacun convient que cette trace graphique est la marque d'un effacement, d'une suppression qui donne lieu ou non à un repentir. De la même manière, que peut-on dire d'un ajout si ce n'est qu'au premier abord, comme pour la suppression, il s'agit en premier lieu d'une marque de rupture dans le flux de la linéarité scripturale d'un texte et d'une marque d'expansion ? Si le trait de barrure est dans quatre-vingt-dix-neuf pour cent des cas la trace incontestable d'un mot que le scripteur a provisoirement abandonné, il n'existe pas toujours de traces aussi graphiquement pertinentes pour l'ajout. Parce que tout peut être ajout, chaque mot écrit l'un derrière l'autre peut être considéré comme le mot qui en remplace un autre et que l'on aurait tendance à réduire à une substitution. Ces deux opérations qui semblent aller de soi dès que l'on observe le processus chronologique et scriptural dans un manuscrit sont en réalité des *universaux* extrêmement difficiles à décrire. Différents niveaux d'entrée dans ces concepts sont nécessaires : ceux qui relèvent de la trace (le trait de biffure pour la suppression ou le signe graphique d'insertion d'un ajout interlinéaire par exemple) et ceux qui relèvent de la qualité³. Cette étude est toujours en chantier dans l'équipe Manuscrit – linguistique – cognition de l'ITEM, tant la distinction poussée à son plus haut niveau entre la trace matérielle et la trace mentale, et donc invisible, est subtile. Si la *TEI* s'appuie essentiellement sur la part visible des traces pour permettre d'attirer l'attention sur une opération de suppression ou d'ajout (« deletion »/« addition »), c'est un moyen de focaliser l'attention du lecteur sur une particularité du processus de révision et de correction dans le manuscrit sans interférer avec le processus interprétatif du chercheur. Ce processus-là (par exemple décider qu'il y a une substitution parce qu'un mot semble remplacer l'autre) se situe à un niveau supérieur et est difficilement codifiable en tant que tel sans argumentation critique. Si finalement peu de concepts sont à repérer en tant que tels⁴ (il n'y a guère que l'ajout, la suppression et sans doute le déplacement qui fassent consensus), la *TEI* en tant que descripteur de ces opérations permet de produire un document compréhensible, transmissible et surtout partageable pour une communauté de chercheurs.

Il y a encore à faire pour concevoir des algorithmes qui permettront la mise en relation des entités scriptographiques avec l'image, en terme d'extraction de « traits » textuels ou graphiques (lignes de texte), de reconnaissance (mappage) de régions et d'indexation globale. Mais il faut surtout approfondir le travail de conception globale d'une édition qui autorisera différents niveaux d'accès aux informations : permettre aux lecteurs d'une œuvre de découvrir son processus créatif tout en autorisant les érudits à l'annoter et contribuer ainsi à créer des échanges entre un statut d'archive et le statut originel du manuscrit en tant qu'œuvre en devenir.

Le numérique est, par sa technologie, davantage adapté à ce type d'édition car il ne suppose pas de refaire un objet à chaque nouvelle contribution. Reste à comprendre comment une telle édition pourra s'inscrire dans un temps historique, se constituer en référence (pour l'imprimé, la date de publication fait référence parce qu'elle a stabilisé l'œuvre critique dans le temps) quand on s'attend à ce que de nouvelles contributions l'actualisent indéfiniment. Doit-on considérer une édition génétique comme « une plateforme d'enregistrements de travaux de recherche » qui trouve sa forme notamment par l'actualisation des mouvements de révision, d'enrichissement, d'hybridation, d'hyperconnexion avec d'autres documents ou productions critiques ?

S'il est encore difficile de répondre à cette question, élaborer un tel type d'édition nous oblige en tout cas à nous interroger sur les changements de statut qu'implique la transposition du manuscrit vers un support immatériel transformant radicalement nos approches.

3. Pour la suppression, on renverra le lecteur aux plus de mille interprétations des traces de suppression de Pierre-Marc de Biasi (*La Génétique des textes*, Paris, Nathan, coll. « 128 », 2000) ; pour l'ajout, il s'agit de ce qui se surajoute entre deux limites syntaxiques (ABC devient ADBC), ce qui est expansion syntaxique, ce qui provient de la marge du feuillet ou qui fait marque d'insertion (le becquet). Quid de la surcharge ? du déplacement ? de la substitution ? qui ne sont en réalité que des déclinaisons ou des combinaisons de ces deux opérations linguistiques supposées simples.

4. Cela n'empêche pas de leur attribuer toutes sortes de sous-concepts (localisation, chronologie...).

AURÈLE CRASSON, Architecte, ingénieur informatique à l'Institut des Textes et Manuscrits modernes (ITEM-CNRS-ENS) ; travaille sur l'édition électronique et génétique de manuscrits d'auteurs. Elle a publié en particulier « L'œuvre d'Edmond Jabès peut-elle se lire sous forme de 0 et de 1 ? », dans *Portraits d'Edmond Jabès* (BnF/Éditions du Seuil, 2000), « Genèse et Hypertextes : échange de partitions » (*Diogenes*, n° 196, 2001), « Du corpus à la région : représentation multi-échelle pour les manuscrits modernes » (*Linguistica computazionale*, 2004), « Récit d'Edmond Jabès, les cinq états du manuscrit » (*Textuel*, 2005), *L'Édition du manuscrit. De l'archive de création au scriptorium électronique* (Academia-Bruylant, 2008).

Aurèle Crasson, Aurele.Crasson@ens.fr